

REVISION SOBRE DIFERENTES METODOS DE OPTIMIZACION BASADOS EN ROUGH SET THEORY

Esteban Alexis Salazar Salazar

Ingeniería de Sistemas

Delio Augusto Aristizábal Martínez.

INSTITUTO TECNOLÓGICO METROPOLITANO

Medellín

2016

RESUMEN

En los años recientes se han publicado diversos artículos relacionados con la técnica de optimización llamada: Conjuntos aproximativos o Rough Set Theory (RST) en sus diversos usos y aplicaciones. En este trabajo se realizó una revisión de diferentes artículos publicados entre los años del 2010 al 2016, sobre los diferentes métodos de optimización que utilizan RST, teoría de conjuntos difusos Fuzzy Sets (FS) y teoría de conjuntos suaves Soft Sets (SS). La revisión consideró las técnicas utilizadas, además, en cuáles han sido implementadas y las tendencias donde la metodología será utilizada en futuras investigaciones y publicaciones; haciendo énfasis en la optimización de búsquedas en sus diferentes tipos, mejoras en la obtención de resultados y reducción de atributos o tiempos de respuesta.

La consulta se realizó en bases de datos científicas relacionados con RST, FS y SS, obteniendo 58 artículos base, donde se clasificaron y agruparon según la técnica utilizada.

Se establece que el RST es una metodología bastante utilizada en diferentes áreas y procesos, validando que es una técnica útil para diferentes aplicaciones como la toma de decisiones, la minería de datos y predicciones, entre otros. Además, se encontró que es un tópico que está atrayendo la atención en diversas investigaciones como también que el RST asociado con algoritmos basados en el comportamiento de la naturaleza (BioMetainspirados), está tornándose en una gran tendencia, abriendo el campo de acción en alternativas de investigación con respecto a la optimización. De la información recolectada, se estableció un análisis comparativo de uso entre las diferentes técnicas encontradas y que interactúan con RST. A su vez, se resalta la capacidad de la teoría y su versatilidad para combinarse con diferentes técnicas y así aplicarse e implementarse en diversos procesos de optimización como se observará en el presente documento.

Palabras Claves— Optimización, Teoría de conjuntos aproximativos, RST, RSFS y RSSS

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

RECONOCIMIENTOS

Siendo los cimientos para formar la persona que actualmente soy, este trabajo que presento para culminar mi carrera y recibirme como profesional va dedicado a mis Padres Gonzalo y Gladys quienes con sus enseñanzas, apoyo constante y sacrificio me motivaron pese a las adversidades a continuar con mis estudios.

Agradezco a mi compañera, amiga y esposa Yese, quien es mi motor para cada día mejorar como persona y profesional.

A mi asesor Delio Aristizábal, a quien no solo agradezco el apoyo y asesoría con este trabajo, sino también por ser el gran pedagogo que tuve hace algunos años y quien me motivo a mejorar mis habilidades.

Y por último al ITM quien me brindo las herramientas necesarias, para convertirme en profesional, por brindarme los espacios para conocer grandes personas y profesionales, me siento orgulloso de pertenecer a esta institución, soy un representante en los diferentes lugares donde he estado, en los cuales mis logros también son los logros de la institución y en los cuales me siento orgulloso de indicar que los realice en ese maravilloso lugar. Gracias ITM.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

ACRÓNIMOS

RST	Rough Set Theory (Teoría de conjuntos aproximativos)
FZ	Fuzzy Set (Conjuntos difusos)
SS	Soft Set (Conjuntos suaves)
ACO	Ant Colony Optimization (Algoritmo de colonia de hormigas)
VPRS	Variable Precision Rough Sets (Variables de precisión en conjuntos aproximativos)
GA	Genetic Algorithms (Algoritmos genéticos)
PSO	Particle Swarm Optimization (Optimización por enjambre de partículas)
ABC	Algorithm Bee Colony (Algoritmo de colonia de abejas)
TS	Tabu Search (Búsqueda Tabú)
GWO	Grey Wolf Optimization (Optimización de lobo gris)
AFSA	Artificial Fish Swarm Algorithm (Algoritmo de cardumen de peces artificial)
FA	Firefly Algorithm (Algoritmo de luciérnaga)
FPA	Flower Pollination Algorithm (Algoritmo de polinización de flor)

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

TABLA DE CONTENIDO

1. INTRODUCCIÓN	5
2. MARCO TEÓRICO.....	8
3. METODOLOGÍA.....	10
4. RESULTADOS Y DISCUSIÓN.....	12
5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO	15
REFERENCIAS	16
APÉNDICE.....	17

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

1. INTRODUCCIÓN

La teoría de los conjuntos aproximativos que viene de la denominación inglesa Rough Set Theory (Pawlak 1982). Es una teoría que se ha venido desarrollando desde comienzos de los años 80. Es una herramienta que es usada para la minería de datos, actualmente tiene aplicaciones en diferentes campos, sobretodo en sistemas de apoyo a la decisión y sistemas gerenciales de información; Desde su publicación hasta la actualidad han surgido diferentes teorías, todas se basan en procedimientos y técnicas que buscan extraer patrones dentro de un conjunto de datos. El objetivo principal de la teoría es lograr que según lo recolectado se pueda indiscernir, es decir que no se pueda distinguir una cosa de la otra, por medio de los sentidos o de la Inteligencia humana [1].

La obtención de resultados claros y concretos es una parte importante en la toma de decisiones, sea de una empresa para comenzar o seguir con un proyecto o en el ámbito personal para saber cuál es la mejor opción a tomar. RST es una herramienta interesante que tiene la característica de obtener y agrupar información que para el humano no es fácil de distinguir, que puede ayudar a optimizar un proceso basado en los resultados que se obtengan o identificar un error y a su vez proceder con una posibilidad de mejora. Si bien los desarrollos e investigaciones brindados por otros autores cuando se creó esta metodología fueron enfocados en el diagnóstico médico para la identificación de enfermedades, para que se pueda realizar un correcto tratamiento al paciente, con este trabajo se quería identificar, si la metodología acá mencionada ha sido implementada en diferentes áreas y procesos, cuáles han sido las tendencias o características que se pueden obtener en las investigaciones publicadas los últimos años y que alcance puede tener esta técnica.

También se pretendió en este proyecto entender los actuales tipos de teoría planteados basándose en la optimización, entendiendo que después del análisis la consecuencia es

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

importante, pero a su vez en cuanto tiempo se pueda obtener; se convierte en un factor influyente basados que, para ser competitivos en el mundo actual, el tiempo que tome una respuesta puede ser la diferencia entre el éxito y el fracaso.

Como objetivo principal se planteó recolectar información sobre los diferentes usos de RST, FS y SS en publicaciones científicas realizadas en los últimos 5 años; Como estas son usadas para la optimización en la búsqueda de atributos o resultados según las aplicaciones para las cuales fueron usadas.

Basado en la información recolectada se procedió a realizar un análisis de las investigaciones, donde se plasmó una clasificación y agrupación sobre los diferentes métodos que son divulgados en los artículos encontrados.

Identificar cada uno de los artículos obtenidos para después proceder a clasificar a que método pertenecen.

Analizar cada una de estas teorías basándonos en la optimización para la obtención de resultados.

Obtener información relevante de cada una de las modalidades que puedan aportar a los argumentos del estado de arte que se quiere realizar.

En el escrito se realizó una clasificación de artículos según el método de optimización usado, esta información se agrupo y se graficó en un diagrama de Venn, después fueron expuestos cada una de las clasificaciones obtenidas, indicando sus aplicaciones y usos relevantes. Por último, se indicaron las tendencias según la investigación realizada.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

2. MARCO TEÓRICO

El RST fue propuesto por Pawlak en 1982 [1]. La idea con la que se basa los conjuntos aproximados se fundamenta en la suposición que alguna información es asociada con cada objeto del universo de discurso. Entre las ventajas de la teoría para el análisis de datos, se indica que esta solo se basa en los datos originales y no necesita cualquier información externa; no es necesaria ninguna suposición acerca de los datos; así, como que sirve para analizar tanto rasgos cualitativos como cuantitativos.

Al desarrollar diversas técnicas para el análisis de datos basados en los conjuntos aproximados, la información es representada por una tabla donde cada fila representa un objeto y cada columna representa un rasgo. Esta tabla se denomina Sistema de Información; más formalmente, donde U es un conjunto finito no vacío de objetos llamado Universo y A es un conjunto finito no vacío de atributos.

Todos los métodos de selección de rasgos contienen dos componentes importantes: una función de evaluación y un algoritmo de generación de subconjuntos. La función de evaluación da una medida de calidad (bondad) de un subconjunto producido por algún procedimiento de generación. Un subconjunto siempre es óptimo con respecto a una determinada función de evaluación. Hay varias categorías de funciones de evaluación, tales como medidas de distancia, las medidas de información (por ejemplo, Entropía), las medidas de dependencia, las medidas de consistencia, y la medida por ciento del error del clasificador. [2]

El algoritmo que se utiliza para la generación de un subconjunto que es donde se indican los parámetros que después pertenecerán al sistema de información se basan en un método de búsqueda. El éxito de la teoría depende que las estrategias de exploración sean bien utilizadas ya que son importantes en el proceso de selección de rasgos que puede

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

consumir mucho tiempo y una búsqueda exhaustiva del subconjunto “óptimo” no es práctico, incluso para los problemas de tamaño moderado.

El uso de reductos (conjunto mínimo de atributos que preserva la partición del universo) en la selección de rasgos o atributos, como también se pueden encontrar en algunos artículos indicados por algunos autores, han sido objeto de investigación en varios documentos, ya que es el comienzo de obtener un resultado lo más aproximado posible, es decir en algunos casos una décima puede ser una diferencia importante según la aplicación y el campo donde se esté aplicando y uno de los objetivos claros de la teoría es que su resultado sea lo más cercano posible de lo que se quiere obtener.

La selección de rasgos es útil en diferentes tareas computacionales, por ejemplo, en el proceso de aprendizaje automatizado. Una apropiada representación del espacio de aprendizaje por la selección de atributos relevantes es un asunto crucial en el aprendizaje. Por lo general, no todos los rasgos que describen el ejemplo son relevantes en el proceso de clasificación y algunos de ellos son irrelevantes. Rasgos irrelevantes aumentan la complejidad de proceso de aprendizaje y disminuyen la calidad del conocimiento inducido.

Los conceptos básicos RST son los conceptos de aproximación inferior y superior. Una definición clásica de aproximación inferior y superior fue originalmente introducida con referencia a una relación de inseparabilidad (indiscernibility relation) la cual es una relación de equivalencia.

Otra de las teorías que nació a partir del primer artículo publicado por Pawlak es FS [3] publicada por Dubois y Prade, esta ha sido aplicada como teoría de la ingeniería de la fiabilidad con un gran éxito hace más de dos décadas, la incorporación del concepto de la teoría de FS dentro del área multidisciplinaria de la teoría de fiabilidad ha estado basada en la modificación de las conjeturas básicas [4].

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

3. METODOLOGÍA

Se procedió a realizar una consulta y recolección de artículos publicados bases de datos científicas en los cuales el contenido expusiera como tema principal las teorías de RST, FS y SS con respecto al uso en los diferentes métodos para la optimización, en la búsqueda de atributos y resultados que se han divulgado en los últimos años, estos trabajos fueron buscados por medio de motores de bases de datos como google académico, ELSEVIER, SPRINGER, revistas de investigaciones científicas enfocadas en tecnología.

Con la cantidad de artículos recolectados, se procedió a analizar el contenido de pertinencia para clasificarlo o no tenerlo en cuenta, teniendo en cuenta como parámetro, los algoritmos usados para la optimización. Se utilizaron herramientas como OneNote para obtener las citas bibliográficas o tomar párrafos importantes que se requerían para la documentación que se necesitaba realizar.

Partiendo de la premisa antes mencionada, se procedió a realizar las búsquedas en el motor de base de datos google académico y diversas bases de datos científicas, realizando consultas con palabras claves tales como: rough set, optimization, performance entre otros, realizándolo en inglés; ya que muchas consultas realizadas en español arrojó muy pocos documentos. Cuando se encontraba información al respecto, se confirmaba que esta publicación perteneciera al rango de fechas planteado en los objetivos, de no ser así este se descartaba.

Otro método de búsqueda fueron las citas bibliográficas, se consultaba las bibliografías en los diferentes documentos seleccionados, de manera tal que se pudiera identificar si se tenía documentos que ayudaran con la consulta.

Cuando se obtuvo una cantidad significativa de documentos, se procedió a realizar una clasificación con el objetivo de entender, hacia donde estaba dirigiéndose la revisión con

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

base a los documentos recolectados. En algunos casos ayudo a entender que se debía realizar más énfasis en algunas búsquedas, ya sea de técnica o fusiones en específico.

Con los artículos seleccionados se procedió con la redacción del documento, se realizó una clasificación acorde a la técnica o técnicas que hace referencia, se identificaba si pertenece a una o más técnicas habían sido usadas dentro de la investigación, siempre validando que en las conclusiones se identificara que el método usado presentara mejoras en los resultados, en una cantidad importante de publicaciones se encontró que diferentes documentos tenían más de una técnica expuesta. Con el objeto de darle facilidad y claridad a la lectura del documento, se procedió a crear un diagrama de Venn, que gráficamente pudiera expresar el estado real de los artículos, en diferentes momentos de la recolección se encontró que había métodos en los cuales se necesitaba más investigaciones para complementar el trabajo.

Una vez conformado el diagrama, se procede a complementar la redacción del artículo, donde estaría separado por técnica expuesta, se establece un orden de importancia teniendo en cuenta primero las mejoras encontradas respecto al RST, después con las fusiones entre algoritmos que han sido más utilizados, se sigue con la explicación de FS y SS y por último se agrega las tendencias.

Se confirma que el diagrama de Venn, fue de vital importancia, por cuanto se podía visualizar e identificar las técnicas más usadas y como se encontraban respecto a otros métodos que apenas estaban comenzando a implementarse como nuevas tendencias.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

4. RESULTADOS Y DISCUSIÓN

RST ha tenido diferentes publicaciones en los últimos años, confirmando que es una metodología que tiene un campo de acción de investigación muy amplio. A medida que el RST ha sido instaurado en diferentes procesos, hace que tenga más difusión y que se presente la necesidad de mejora, ya que, en el mundo actual, el tiempo que tome una respuesta puede ser la diferencia entre el éxito y el fracaso. Las mejores implementadas en la teoría y las fusiones que esta ha tenido con diferentes técnicas han demostrado que se pueden obtener mejores resultados para mejorar los rendimientos en las diferentes aplicaciones en las cuales ha sido asentado.

RST ha demostrado que puede fusionarse en diferentes métodos de optimización, inclusive se encuentra que existen aplicaciones donde hay hasta tres técnicas simultáneamente, confirmando la versatilidad que tiene la teoría para poder ser aplicada en diferentes usos.

Una de las tendencias que se identifica en las técnicas fusionadas con RST, son los algoritmos BioMetainspirados, los cuales han impactado positivamente el rendimiento de la metodología, demostrando que son una buena alternativa para obtener mejores resultados en un tiempo óptimo.

Con los 58 artículos seleccionados se obtuvo una clasificación de uso de las diferentes técnicas con respecto a RST, el cual se puede observar en la tabla 1.

Técnica Usada	Artículos donde fue usado	% Promedio
Ant (Colonia de hormigas)	9	15.5
Bee (Colonia de Abejas)	4	6.9
Firefly (Luciérnaga)	2	3.4

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Fish Swarm (Cardumen de Peces)	2	3.4
Flower P (Polinización de flor)	1	1.7
Fuzzy Set (Conjuntos difusos)	8	13.8
Genetic Alg (Algoritmos genéticos)	6	10.3
Grey Wolf (Lobo Gris)	2	3.4
Rough Set (Conjuntos Aproximativos)	7	12.1
Set tree (Árbol)	8	13.8
Soft Set (Conjuntos suaves)	2	3.4
Swarm (Partículas)	14	24.1
Tabu	3	5.2

Tabla 1. Tabla de porcentajes por artículo de Revisión sobre diferentes métodos de optimización basados en Rough Set Theory.

En la tabla 1, se identifica que el algoritmo de partículas, es el más utilizado estando en el 24% de los artículos recolectados, también se identifica que es muy usado para fusiones con otras técnicas. Los algoritmos de hormigas, abeja, genéticos y de árbol que suman el 46.5%, aunque son de los que llevan más tiempo comparados con otros, confirman que siguen siendo tenidos en cuenta para investigaciones que involucren mejoras.

La teoría de conjuntos difusas que representa el 13.8, indica que es tenida en cuenta por los investigadores y que a su vez comienza a ser usada para fusiones, contrario a la teoría de conjuntos suaves que con el 3.4%, demuestra que todavía hay muy pocos documentos donde se hable de mejoras.

Algoritmos como el de luciérnaga, cardumen de peces, polinización de flor y lobo gris, donde se identificó que son nuevas metodologías y que podrían ser usadas de manera más constante en el futuro hicieron parte del 11.9% de las investigaciones.

En la figura 1 se puede observar la interacción que existe en las diferentes técnicas obtenidas en los 58 artículos seleccionados, cabe resaltar que el RST es quien es la técnica objetivo de referencia con respecto a las demás técnicas que interactúan con ella.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

De las metodologías que se puede ver graficadas en el diagrama, se identifica que el SS, todavía no presenta fusiones con otras técnicas para optimizar sus resultados, se identifica que por el momento las mejoras han sido implementadas sobre el método, pero no hay estudios que relacione alguna unión para la mejora de rendimiento.

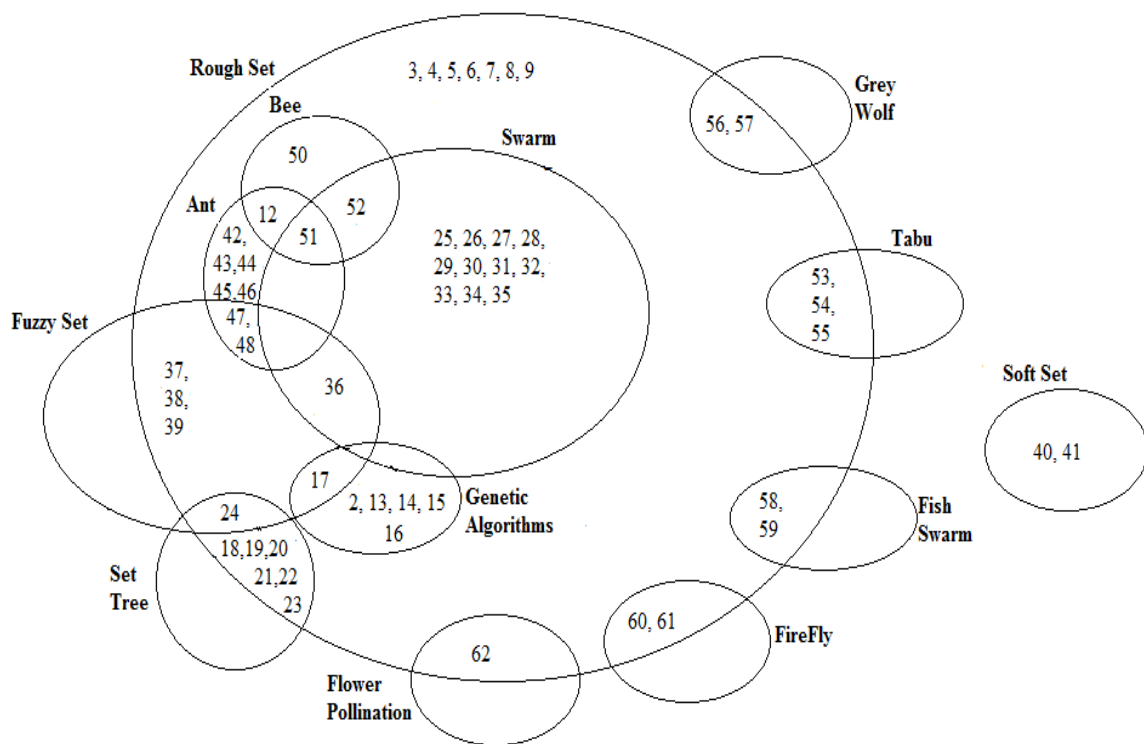


Figura 1. Diagrama de Venn Revisión sobre diferentes métodos de optimización basados en Rough Set Theory.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

5. CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

Inicialmente se planteó recolectar información sobre las diferentes funcionalidades en publicaciones científicas realizadas los últimos años sobre Rough set, Fuzzy set y Soft Set; Como habían sido usadas para la optimización en la búsqueda de atributos o resultados, sin embargo, en el proceso de recolección se puede encontrar que es una metodología muy usada y este proceso se enfocó en artículos escritos en inglés, en el proceso se encontraron muy pocos en español enfocados en las mejoras y no se realizó una investigación sobre artículos escritos en otros idiomas, sin embargo se pudo identificar que se tiene bastante material escrito en esta metodología.

Una de las pautas que marco esta investigación, fueron los algoritmos bioinspirados para fusionarse con la teoría, estos están basados en el comportamiento animal, en el proceso de recolección se identificó que hay diferentes algoritmos diferentes a los expuestos en el documento técnico sin embargo no fueron agregados ya que presentaron ningún tipo de mejora en sus conclusiones, sin embargo, se identificó que este tipo de fusiones son tendencia para los próximos años.

Los algoritmos de partículas, son los más usados con esta teoría, también son los que más permiten fusionarse con otros algoritmos, a pesar que es uno de los métodos que más tiempo llevan, sigue estando a la vanguardia como el más usado por los diferentes investigadores, será interesante en el futuro como seguirá comportándose este algoritmo, si seguirá marcando la pauta como uno de los más usados o si tal vez con la capacidad de fusión que este tiene pueden existir investigaciones donde se exponga mejoras superiores a las encontradas en este trabajo, asumiendo que se vienen con más fuerza las nuevas tendencias, en especial las basados en comportamientos de la naturaleza.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

REFERENCIAS

-
- [1] Z. Pawlak. "Rough Sets". *International journal of Computer and Information Sciences*. Vol. 11. 1982. pp. 341-356.
- [2] M. Kudo, J. Sklansky. "Comparison of algorithms that select features for pattern classifiers." *Pattern Recognition*. Vol. 33. 2000. pp. 25-41.
- [3] D. Dubois y H. Prade "Rough fuzzy sets and fuzzy rough sets" Vol. 17.2007 pp 191-209
- [4] V. Ravi, P. J. Reddy, y H.J. Zimmermann "Fuzzy Global Optimization of Complex System" *Fuzzy Systems* Vol. 8. 2000. Pp. 241-248.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

APENDICE

Artículo titulado “Revisión sobre diferentes métodos de optimización basados con Rough Set Theory”.

*Esteban Alexis Salazar-Salazar¹, Delio Augusto Aristizábal-Martínez²,
IInstituto Tecnológico Metropolitano - ITM, Calle 73 # 76 a 354, Medellín, Colombia
correo-e: estebansalazar67256@correo.itm.edu.co¹, delioaristizabal@itm.edu.co²*

Resumen- En los últimos años se han publicado diversas investigaciones relacionados con Rough Set Theory (RST) en sus diferentes usos y aplicaciones. En este documento se realiza una revisión de artículos publicados entre los años del 2010 al 2016, sobre diferentes métodos de optimización que utilizan RST, Fuzzy Sets (FS) y Soft sets (SS). La revisión considera las tendencias y hacia dónde se dirigen las futuras publicaciones o estudios que utilizan esta metodología, haciendo énfasis en la optimización de búsquedas en sus diferentes tipos, mejoras en la obtención de resultados, reducción de atributos o tiempos de respuesta. Las búsquedas exhaustivas no son lo suficientemente efectivas en cuanto a recursos y tiempos de respuesta, lo cual ha obligado a la implementación y uso de nuevas tendencias que se pretenden mostrar en el presente documento. Se necesita una combinación de agilidad y obtención de datos concretos para la mejora de resultados o técnicas, por tal razón cuando se realiza una optimización sobre un proceso, se aplica una modificación sobre el método o fórmula usual de proceder, con el principal objetivo de obtener resultados que estén por encima de lo esperado y así puedan ser implementados en diferentes campos de acción.

Palabras Claves— Artículos publicados, Optimización, Teoría de Conjuntos aproximados, Revisión de Métodos de Optimización

Abstract -The last years, several articles related with Rough Set Theory (RST) about their different applications and uses have been brought out. In this paper we made a summary of all these documents published from 2010 to 2016, about different optimization methodologies who use RST, Fuzzy Sets (FS) and Soft sets (SS). This revision considers the tendencies, where will be point the future publications or studies who uses this methodology, making an emphasis in the

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

searching optimizations on their different types, improvements to get results, attributes reductions or response times. The exhaustive searches are not practical enough about resources or response times, this have been forced to deploy and use new tendencies who will be indicated on this document. It is necessary and agility combination and concrete get results to improve the results or techniques, for this reason when an optimization is made in a process, a modification was applying about a method or usual formula to proceed, with the principal goal to obtain the results who are higher than expected and with it could deploy on different fields of action.

Palabras Claves— Optimization, Optimization methods revision, Published articles, Rough set theory.

1. INTRODUCCION

La teoría de los conjuntos aproximativos que viene de la denominación inglesa Rough Set Theory (RST) desarrollada por Zdzislaw Pawlak en 1982 [1]. Es una herramienta usada para la minería de datos, la cual tiene aplicaciones en diferentes campos tales como, implementación para predecir la deserción académica superior, esperanza de vida de los pacientes con enfermedades terminales y tratamientos a seguir, dependiendo sus síntomas, también ha sido utilizada para solucionar problemas de reducción de datos, predicciones de quiebras empresariales, sistemas de evaluación de decisiones, entre otros [2], [3], [4], [5], [6], [7], [8], [9]

Desde su publicación han surgido diferentes teorías, todas se basan en procedimientos y técnicas que buscan extraer patrones dentro de un conjunto de datos. El objetivo principal de la teoría es lograr que según lo recolectado se pueda indiscernir, es decir que no se

pueda distinguir una cosa de la otra, por medio de los sentidos o de la inteligencia humana [1]. Basados en esto, se puede encontrar objetos o atributos (acciones, alternativas, medidas, estudiantes, pacientes, etc.) que cuando se obtienen, hacen parte de un mismo tipo de información el cual se puede agrupar, esto quiere decir que se puede indicar que esta información es indiscernible.

Varios autores se han basado en lo desarrollado por Pawlak años atrás y por medio de sus investigaciones y análisis han encontrado diferentes maneras de ampliar la teoría, es decir, han encontrado que combinándola con diferentes algoritmos se han obtenido mejores resultados en la reducción de datos o en la optimización al momento de obtener deducciones. Como a su vez otros han tomado conceptos y se basan en ésta y fusionándolos con sus ideas desarrollan otras teorías como la teoría de aproximación de conjuntos difusos o FS [10], otra teoría que será también planteada en este documento la cual se

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

llama Teoría de Conjuntos Suaves o SS [11], que utiliza la parametrización de subconjuntos en un conjunto. Al momento de realizar una optimización se realiza una modificación donde se requiere tener un rendimiento que mejore lo antes planteado; en este documento se plasmará la exploración realizada sobre las diferentes combinaciones usadas en las cuales se han podido identificar que las teorías han sido fusionadas con algoritmos metaheurísticos, usando los parámetros dados por el usuario sobre unos procedimientos genéricos y abstractos de una manera que se espera sea eficiente, se indicaran cuales han sido más consultadas en los últimos 6 años y cuales según lo recolectado marcaran tendencia en un futuro, los cuales serán sin duda alguna patrones para futuros estudios. Al ser una metodología que está siendo tendencia y que es utilizada en diferentes campos, debe cumplir los requerimientos necesarios para ser competitiva, no solo basta con tener resultados exactos, se necesita también tiempos de respuesta ágiles, que en algunos casos según la aplicación usada sean competitivos, este último es uno de los objetivos de esta revisión la cual desea apoyar al lector para que tenga un panorama de una herramienta ágil y constante, que según su necesidad pueda escoger las diferentes metodologías que serán acá expuestas.

2. METODOLOGIA

La metodología utilizada para dicha investigación, se basó en 58 artículos publicados en revistas científicas entre los años de 2010 a 2016. El primer parámetro de búsqueda fue consultar diferentes publicaciones que tuvieran relación con publicaciones que indicaran optimizaciones en RST y las diferentes mutaciones que se han venido publicando y marcando tendencia o uniendo la teoría con diferentes algoritmos, tales como el de la Colonia de Hormigas Ant Colony Optimization (ACO por sus siglas en inglés Ant Colony Optimization). Por otro lado, en el proceso de recolección se encontraron nuevos algoritmos que se unieron al concepto, por ende, se procedió a realizar un análisis de estos y consultar nuevas tendencias que sustenten esta averiguación.

De manera conjunta, también se realizó la misma labor buscando mejoras en FS y SS.

En el diagrama de Venn de la Figura 1, indica la cantidad de artículos que fueron consultados y luego clasificados, los números que se visualizan en esta imagen hacen alusión a la referencia del artículo que se puede encontrar en la bibliografía de este documento, como un ejemplo el artículo citado con número [12], después de haber sido analizado, se ubicó en la intersección del conjunto Bee y el conjunto Ant los cuales hacen alusión a los diferentes tipos de técnicas que fueron agrupadas después de la recolección de los escritos

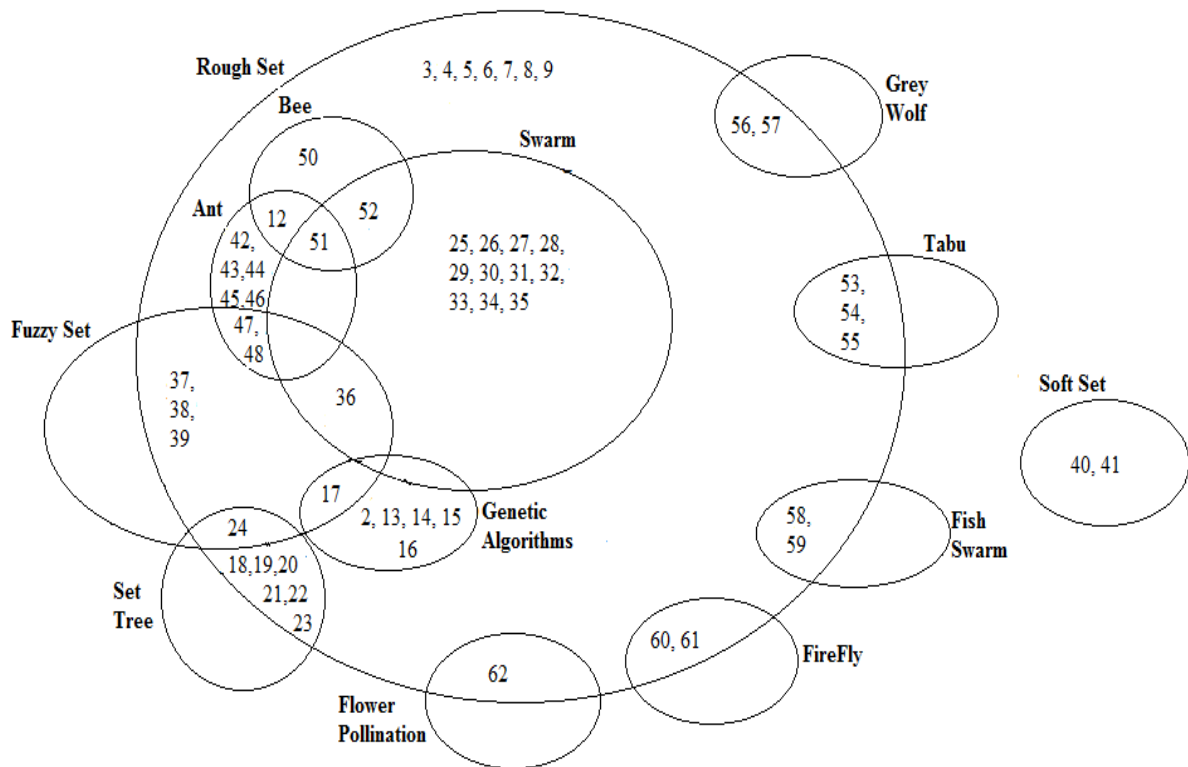


Figura 1. Diagrama de Venn Revisión sobre diferentes métodos de optimización basados en Rough Set Theory.

3. OPTIMIZACIONES EN RST

En cualquier proceso sin importar su aplicación, cuando se desea realizar una mejora se procede analizar que parte afectar o en algunos otros, según algo planteado se comienza con el proceso de análisis, en esta revisión se puede encontrar que hay diferentes maneras en las cuales los autores han tomado como base para sus mejoras diferentes métodos, uno de estos son las aproximaciones. La teoría se basa en aproximar cualquier concepto a un subconjunto complejo del dominio, un ejemplo es una clase en un problema de

clasificación supervisada, por un par de conjuntos exactos, llamados aproximación inferior y aproximación superior del concepto [1], sin embargo, en uno de los artículos más aportantes a otros estudios y más citados los últimos años, se creó un concepto llamado aproximación positiva, que se basa en acelerar el proceso heurístico en diferentes algoritmos de los atributos de reducción para la mejora de procesos en equipos de cómputo [7]. Luego con la llegada de datos en la nube, la inteligencia de negocios y la minería de datos, Dean y Ghemawat de Google, crearon un

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

concepto llamado MapReduce que luego fue tomado para crear un concepto paralelo basado en sus algoritmos y combinado con RST mostraron grandes resultados al momento de obtener datos [4]. Siguiendo por la rama de los trabajos que se enfocaron en las aproximaciones, en 2012 se mejoró un concepto simple y fácil de implementar basado en RST llamado vecindario, que tiene como base, técnicas y clasificaciones para los atributos de reducción, este se usa para procesamiento de datos numéricos, en esta publicación se creó un algoritmo simple para actualizar las aproximaciones (NAUANRS) y dos algoritmos más rápidos que también actualizan (FAUANRS-MI y FAUANRS-ME), cuando los múltiples objetos entren o salgan de una tabla de decisión llamada vecindario [5]. Uno de los patrones que se pudo encontrar en esta investigación, fue los pocos documentos escritos por autores castellano, sin embargo, se encontró uno que aplica bastante a este estudio de la Universidad de Antioquia, donde presentan los conceptos y medidas fundamentales de RST, se analiza la extensión de la misma para el caso de rasgos con dominios continuos y se muestra cómo el empleo de relaciones de la similitud en RST lleva a la necesidad de calcular los pesos de los atributos [3].

Como se indicó anteriormente una las particularidades de los conjuntos aproximados, es poder implementarse a diferentes aplicaciones, un ejemplo fue un modelo mejorado para el uso de la tierra y la recuperación del uso de esta en un inventario para un pantano en China,

utilizando las variables de precisión de RST (Variable Precision Rough Sets o sus siglas en inglés VPRS) teniéndose en cuenta para futuros estudios ecológicos [8].

La reducción de Costos por medio de las mejoras en los atributos de reducción es otro método planteado para la optimización del concepto, el atributo de reducción es una interpretación de cómo encontrar el conjunto de atributos mínimos que puede preservar o mejorar uno o diferentes criterios para realizar la decisión del costo siendo este mas intuitivo y razonable, en estos dos documentos se encontraron estudios enfocados en la optimización de este método [6], [9].

3.1 OPTIMIZACION ROUGH SET Y ALGORITMOS GENÉTICOS

Una vez el RST estaba establecido y se convirtió en un método bastante usado, comenzaron a fusionarse con otros conceptos, uno de estos y sin duda uno de los más usados, son los algoritmos genéticos (GA por sus siglas en inglés Genetic Algorithms), los conjuntos de aproximación de baja y alta en un clúster, pueden describir mejor el objetivo necesitado, con esta base un método de agrupamiento basado en RST y GA puede ser provisto, los resultados experimentales realizados muestran que el algoritmo puede converger en una óptima solución global que puede ser usado en cualquier modelo y obtener

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

mejores resultados [13]. Uno de estos usos se realizó en un modelo híbrido escrito en 2010, donde su objetivo era predecir los precios de las acciones en una bolsa de valores, demostrando que los resultados comparando el modelo RST, luego el GA utilizado para un retorno de valores y por último se tomó el concepto híbrido, donde se demostró su mejora notable, confirman los buenos resultados de la fusión [14]. El RST y los algoritmos genéticos se implementan para generar reglas de decisión que caracterizan los estudiantes en riesgo de deserción, con el fin de ayudar a la construcción de metodologías y herramientas que permitan a las instituciones incrementar la retención de los estudiantes [2], estudios similares han podido arrojar con un 98.3% de eficiencia usando metodología geométrica con el cual se realiza el análisis de información con el algoritmo genético integrado en ROSETTA [15].

Existe un modelo llamado RPGA (Rough Penalty Genetic Algorithms), que tiene los siguientes objetivos, ajustar los coeficientes de acuerdo a sus restricciones y analizar las restricciones ineficientes por medio del RST, con la ventaja que puede ser aplicada automáticamente a coeficientes, demostrando que no solo puede encontrar soluciones cercanas muy optimas, si no también obtener resultados robustos para funciones lineales y funciones de restricciones no lineales [16].

3.1.1 OPTIMIZACIONES RST, FS Y ALGORITMOS GENETICOS

En el proceso de consulta se encontró que hay publicaciones donde se tomaron conceptos de más de un tipo de algoritmo, como el que se indica en este caso donde se realizó una fusión de RST, FS y GA, desarrollado para conducir la búsqueda de conjuntos adecuados para las instancias, indicando resultados de mejoras considerables [17].

3.2 OPTIMIZACIONES RST Y ALGORITMOS DE ARBOL

Otro de los modelos más usados con RST son los Algoritmos de Árbol en sus diferentes tipos, uno de ellos son los de decisión, es un algoritmo de aprendizaje inestable, se representa como un diagrama de flujos donde cada nodo interno denota una prueba o un atributo, cada rama representa una salida de la prueba y cada nodo hoja (o nodo terminal) contiene una clase, dada su popularidad un número de algoritmos de conjuntos efectivos han sido creados esta última década, siendo muy utilizado para la minería de datos que cada día se convierte en una tendencia filtrando la información redundante, hay estudios de estos modelos para bolsas de valores [18] o sistemas de detección de intrusos (IDS) que se han convertido en componentes esenciales de seguridad informática [19]. También hay mejoras implementadas para la clasificación de textos, que es una clasificación automática de documentos

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

dentro de categorías predefinidas y que son una parte importante en la comunidad Web de minería de datos [20], también se encontraron mejoras en algoritmos de decisión de árbol basados en VPRS usando grados de dependencia para la división de atributos [21].

En la idea de investigar algoritmos más ágiles y efectivos que los comúnmente usados, se estructuró uno llamado Power Set Tree (PST) creado en 2011 y que se basa en la idea anterior del algoritmo de árbol, pero con mejoras en las ejecuciones [22]. Muchos estudios de RST han sido planteados para el área médica y en búsqueda de optimizaciones está no es la excepción, como este documento planteado para la reducción de tiempos en búsquedas de tumores cerebrales fusionado con PST [23].

3.2.1 OPTIMIZACIONES RST, FS Y ALGORITMOS DE ARBOL

FS tampoco ha estado ajeno a fusiones con otros algoritmos, uno de los estudios identificados y que ha comprobado mejoras importantes, fue realizado con algoritmos de Árbol, el cual concluyó que las reglas de clasificación de la teoría convertidas en el árbol de decisión fuzzy propuesto en este estudio, demostrando una habilidad de generalización más fuerte [24].

3.3. OPTIMIZACIONES RST Y ALGORITMOS DE ENJAMBRE DE PARTICULAS

El algoritmo de enjambre de partículas (PSO por sus siglas en inglés Particle Swarm Optimization) es el más utilizado para combinar con RST y FS. Está basado en el comportamiento social de algunos seres vivos tales como hormigas, abejas, murciélagos, luciérnagas y peces, se encontró un artículo el cual realizó una comparación sobre 6 algoritmos diferentes donde se expresa sus ventajas, desventajas, aplicaciones usadas y metodologías empleadas, siendo uno de los más completos encontrados en esta recolección [25]. No todas las técnicas usadas para realizar mejoras a un proceso garantizan una optimización, sin embargo, algunos métodos usados proveen unas prometedoras maneras de mecanismos de selección de atributos para la implementación y mejora de minería de datos, tomándolo como uno de los ejemplos más usados, en los siguientes documentos se propusieron diferentes técnicas basadas en PSO y RST [26], [27], [28].

La idea de selección de atributos de un subconjunto es preservar el conocimiento dado o decisión del sistema, hay muchos algoritmos de selección de características, alguno de ellos, son filtros simples aplicados sobre cada atributo de manera separada, uno de los métodos para encontrar un buen subconjunto de atributos es calcular reducciones, estos pueden ser divididos en dos grupos, filtros y envolturas [29], [30]. La idea de reducir tiempos ha alentado muchas investigaciones en el estudio de efectividad en RST en varias aplicaciones,

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

incluyendo medicina, farmacología, sistemas de control, diagnósticos de faltas, categorización de textos, ciencias sociales, circuitos de comunicación, predicciones económico, financieras, proceso de imágenes etc. [31].

Siguiendo la línea de PSO, en una de las publicaciones se usó un algoritmo de búsqueda para mejorar los tiempos de respuesta, el cual fue comparado con 10 repositorios de datos brindado buenos resultados [32].

Los algoritmos de detección de valores atípicos son usados en aplicaciones de minera de datos, tales como detección de fraudes, detección de intrusos en redes, entre otros. En este documento se realizó un algoritmo que procediera a encontrar estos datos de manera óptima [33].

Basados en el concepto de vecindario antes anunciado el cual ha sido implementado para RST, también se tienen estudios de mejora incluyendo el algoritmo PSO, en este caso se implementó para investigaciones en ejercicios gimnásticos [34].

Cerrando los tipos de investigaciones de este tipo, otra problemática que ha sido atacada al momento de pensar en la mejora de un proceso ha sido la extracción de reglas, que han sido utilizadas para reducir el problema de tamaños y el espacio del resultado de la búsqueda. En el diseño de patrones de clasificación, este puede ser mejorado en la calidad y velocidad de esta, la extracción de reglas es un área emergente en la búsqueda de inteligencia de

negocios, se encuentran muchos métodos que han sido creados, sin embargo, en este documento se realizó una fusión de elementos donde prima el algoritmo PSO donde consta de dos fases, la primera es la preparación de datos y la segunda extracción de reglas [35].

3.3.1 OPTIMIZACIONES RST, FS Y PSO

Los algoritmos de enjambre de partículas (PSO) sin duda alguna son los más usados para RST, también se encuentran algunos artículos relacionados con mejoras uniéndose con FS, con la particularidad que se pueden encontrar diferentes reducciones mínimas de datos, en este documento se presenta un concepto de topología anillo y se expresa una nueva regla de actualización de velocidad para obtener eficiencias [36].

3.2 OPTIMIZACIONES RST Y FS

FS es recomendado para manejar atributos reales, discretos o como mezcla de ambos, es una herramienta adecuada para manejar información inexacta o con ruidos, la aplicación más relevante del FS, es encontrar los atributos de los subconjuntos más exactos. Una propuesta llamada FRNN (Fuzzy-rough nearest neighbours) obtuvo mejoras alrededor del 25% en los valores de decisión que estaban corruptos por ruido en sus datos [37]. Uno de los estudios recientes propuso una mejora para FS en la reducción de tiempos, que inclusive fue

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

comparado con 18 algoritmos diferentes y se pudo comprobar su superioridad en diferentes operaciones y aplicaciones, incluso se realizaran futuros estudios con el objetivo de manejar información más ruidosa [38], otro ejemplo de mejoras fue realizado basándose en el uso de matrices con los mismos conceptos de los dos últimos artículos antes mencionados [39].

3.3. OPTIMIZACIONES RST Y SS.

Molodstov, desarrollo el concepto SS, una nueva herramienta matemática para manejar las incertidumbres que son libres de dificultades. En distintas investigaciones Molodstov apunto a diferentes direcciones para las aplicaciones de SS, tales como teorías de juegos, investigación de operaciones y suavidad de funciones. La teoría SS, es una familia parametrizada de subconjuntos del universo, los parámetros son conceptos vagos del mundo real desde el punto de vista de teoría FS [11]. En los años recientes, las investigaciones sobre este campo han aparecido de manera más frecuente, y se tiene un gran progreso de logros en el aspecto de la teoría, al mismo tiempo hay algunos progresos sobre las aplicaciones prácticas de esta metodología.

Una de las investigaciones más completas que se tienen actualmente sobre optimización en esta herramienta, fue realizada en 2010, que ha sido una de las más citadas desde su publicación, en la cual se redefinió las operaciones de Molodstov, para hacerlas más

funcionales, de esta manera se planteó una mejora y se obtuvieron nuevos resultados, también se definió la producción de la teoría y la función (unit) de decisión, realizando un método el cual selecciona un conjunto de elementos óptimos desde las alternativas. Finalmente se presenta un ejemplo que demuestra que el método es exitoso aplicado a diferentes problemas que contienen incertidumbres [40].

En otra investigación se estableció la optimización del modelo para determinar pesos a los atributos, para esta situación especial donde la información sobre los pesos de los atributos es completamente desconocida, se definió otro modelo de mejora. Para solucionar este caso, se desarrolló un formula de extracción simple, la cual puede ser usada para determinar los pesos de los atributos [41].

3.4. OPTIMIZACIONES RST Y COLONIA DE HORMIGAS.

El algoritmo de colonia de hormigas (ACO por sus siglas en inglés Ant Colony Optimization) es metaheurístico, inspirado en el comportamiento de las hormigas quienes no tiene vista y son capaces de encontrar la ruta más corta sin antes validar todos los caminos posibles entre la comida y su nido. Tales como los propuestos en las combinaciones de la información de entropía, donde los experimentos propuestos pueden obtener no solo la más mínima reducción de subconjuntos eficientes y efectivos de completos e incompletos atributos [42].

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

En otros estudios, los métodos combinados de atributos y efectividad en la dependencia para la construcción de un estado de reglas de transición, las cuales mejoran la eficiencia de la búsqueda de ACO, sumando que para obtener características de subconjuntos optimas, el atributo de dependencia basado en RST son usados para la construcción de una estrategia de actualización de feromonas [43]. En el mundo de las aplicaciones reales, el dominio de pocos o de todos los atributos de los conjuntos de datos puede ser continuos, estos atributos pueden ser discretos, como un paso antes de obtener el atributo de reducción, en este documento se propuso un algoritmo enfocado hacia el problema de reducción sobre datos continuos en un RST, el algoritmo propuesto en el año 2015 no necesita información extra o un amplio conocimiento aparte de los conjuntos de datos continuos [44].

El algoritmo ACO ha sido aplicado de manera exitosa en varios problemas complejos como asignaciones cuadráticas, la ruta de vendedores, programaciones, etc. Esto es una atracción particular para la selección de características desde la información no heurística, que puede llevar a la búsqueda de obtener cada vez un subconjunto mínimo de manera óptima. Sin embargo, las hormigas pueden descubrir los mejores rasgos de combinaciones atravesando todos los puntos necesarios para obtener estos datos, en este documento publicado en 2010, se propuso una nueva fusión con RST, mejorando la selección de características

basados en ACO [45], convirtiéndose en una base importante para diferentes publicaciones.

Este algoritmo ha sido implementado en casi todas las aplicaciones de ingeniería como fundiciones de acero, reconciliación de datos y estimación de parámetros en sistemas dinámicos, teorías de juegos, optimización de manejo de combustibles en ingeniería nuclear, problemas de objetivos de rastreo en procesos de señales, diseño de materiales en dispositivos de manipulación, matemáticos, modelos cinéticos, diseño de software, diseño de rutas para vehículos, problemas de asignaciones cuadráticas, problemas de mutación entre otros. Este documento analizo diferentes investigaciones e implementaciones y propone una modificación en el algoritmo el cual fue aplicado para el problema de redes de ruta y fue comparado con algoritmos de ruta tradicionales demostrando resultados de mejora [46].

3.4.1 OPTIMIZACIONES RST, FS Y ACO

Una fusión que ha brindado resultados, es la hecha por FS y ACO, donde se han podido mostrar resultados de mejora con respecto al plan de ruta que debe tener un vendedor para cubrir una mayor área en menor tiempo [47]. A su vez también hay mejoras para la selección de características por medio de una búsqueda en el área de minería de datos, reconocimiento de patrones y

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

comunidades estadísticas. La principal idea de la selección de características es seleccionar un subconjunto de propiedades disponibles para eliminar características con información no predictiva [48].

3.5 OPTIMIZACIONES RST Y COLONIA DE ABEJAS

El algoritmo de colonia de abejas (ABC por sus siglas en inglés de Algorithm Bee Colony) fue propuesto por Karaboga y Basturk en 2007, es un nuevo enfoque metaheurístico, el cual simula el comportamiento de búsqueda de las abejas obreras y su desarrollo original para solucionar problemas de optimización sobre los espacios continuos [49], básicamente hay tres grupos de abejas en el algoritmo ABC, llamadas abejas obreras, observadoras y exploradoras, cada una juega un rol diferente en la búsqueda de nuevos recursos de comida, mientras en una búsqueda local es llevado a cabo por las abejas obreras y observadoras, la búsqueda global es administrada por las observadoras y las exploradoras, así el proceso de exploración y explotación puede ser balanceado.

En la búsqueda de mejoras, diferentes autores han modificado las bases de los algoritmos para conseguir dicho objetivo, uno de los casos es el publicado en 2015, que propone un algoritmo de esquema de búsqueda binario multidimensional, el cual es presentado con mejoras locales en la búsqueda de estrategias para mejorarla

en su diversidad y eficiencia, el algoritmo es comparado contra otros metaheurísticos populares para el problema de obtención de atributos mínimos [50].

3.5.1 OPTIMIZACIONES RST, BEE Y ACO

La reducción de dimensionalidad es vista como un paso pre proceso importante para el reconocimiento de patrones y la minería de datos, el RST considera todos los conjuntos de atributos como un subconjunto completo de búsqueda, comparado con todas las posibles combinaciones de conjuntos de atributos que de por sí ya es complejo. En este estudio se desarrolla una nueva manera mejorada de reducción de atributos de RST Y ABC, el cual encuentra las características de los subconjuntos independientes basados en la decisión de atributos (clases) como primer paso y luego encuentra el reducto final [12].

3.5.2 OPTIMIZACIONES RST, ANT, BEE y PSO

Las investigaciones de optimización son de las más frecuentes en este tipo de teorías y en ocasiones según los pasos se proponen diferentes algoritmos, en los siguientes documentos se hace una fusión de ABC con PSO y ABC, uno de los métodos más usados siempre estará relacionado a los atributos, siendo un proceso importante que debe abundar en el ruido de los datos, irrelevancia o características engañosas. Un método

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

extenso debe usarse con este propósito, en este documento utilizado para el área médica se propone un algoritmo mixto, aplicado para encontrar los atributos mínimos [51].

3.5.3 OPTIMIZACIONES RST, ABC Y PSO

Otra opción de mejora es reducir el número de características que se pueden lograr, la exactitud superior de clasificación usando todo el conjunto de atributos. Para algunos conjuntos de datos, la exactitud ha incrementado significativamente, aunque el número de atributos a reducido también de manera significativa. El método propuesto que combina ABC y PSO, indica mejores resultados para la mayoría de los conjuntos de datos probados comparado con otros algoritmos [52].

3.6 BUQUEDA TABU

La búsqueda de tabú (TS de sus siglas en inglés Tabu Search) fue originalmente propuesto por Glover en 1986, es un método heurístico el cual muestra sus tiempos de respuesta de manera exitosa solucionando diferentes búsquedas, combinadas en algunas exploraciones operacionales de literatura. El sistema combinado con RST usa una representación binaria de soluciones en el proceso de selección de atributos, los resultados numéricos indican que el método propuesto ha sido superior en algunas aplicaciones, guardando costos

computacionales e improvisando la exactitud de algunos modelos de clasificación. La información de búsqueda TS mantiene los datos en caminos que han sido visitados previamente, usando las estructuras de memoria conocidas como lista tabú, esto ha sido aplicado a un rango amplio de optimización de inconvenientes donde están involucrados varias clases de problemas de números enteros.

Un novedoso método de selección de atributos basados en RST y TS ha sido propuesto para un problema de puntuación de crédito, una nueva diversificación e intensificación de elementos ha sido embebida en una fusión llamada FSRT para alcanzar mejores tiempos de respuesta, el experimento numérico fue realizado con un conjunto de datos de crédito de Japón presentando eficiencias con este modelo [53].

En otra publicación se expusieron mejoras con un método dinámico llamado Dynamic Tabu Search (DTSAR) para solucionar inconvenientes de reducción de atributos y así obtener ejecuciones más ágiles [54]. En contraste a optimizaciones convencionales de búsquedas TS, un modelo inteligente híbrido propone un esquema de parqueo con una metodología RST incorporando la entrada de incertidumbre que prima en las soluciones de parqueaderos [55].

3.6. TENDENCIAS DE RST

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

En modelos de investigación existen todo tipo de tendencias, algunas que siguen la rama de desarrollo, como también se pueden identificar otras que salen del tema principal del método, sin embargo en el RST este tema no es indiferente, realizando esta investigación, se ha podido identificar algunos algoritmos que están empezando a marcar tendencia y donde se identifica hacia dónde va el modelo en un futuro cercano, enfocándose precisamente en esas investigaciones que seguramente seguirán siendo usadas y que son fieles a la base de RST. Se ha podido encontrar que los últimos artículos tienen como un patrón importante fusiones con algoritmos metaheurísticos, que son algoritmos aproximados de optimización y búsqueda de propósito general. Son procedimientos iterativos que guían una heurística subordinada combinando de forma inteligente distintos conceptos para explorar y explotar adecuadamente el espacio de búsqueda, siendo estos inspirados por la naturaleza, específicamente por el comportamiento animal. En este documento se han podido explicar algunos como el ACO y ABC, los cuales han sido muy utilizados en los últimos años. A continuación, se expondrán los artículos que se han enfocado en nuevos métodos fusionados con RST y que se han enfocado en la optimización.

3.6.1 TENDENCIA DE RST Y ALGORITMO DE LOBO GRIS

El algoritmo del lobo gris (GWO de sus siglas en inglés Grey Wolf Optimization) asemeja la jerarquía en el liderazgo y el mecanismo de caza de los lobos en la naturaleza, comparado con algoritmos genéticos, el GWO no necesita de operaciones complejas tales como mutaciones o cruces, este necesita solo de operaciones matemáticas fáciles y primitivas, y computacionalmente no consume tanto en términos de memoria y tiempos de ejecución, el GWO tiene capacidades de búsqueda amplias y puede de manera efectiva encontrar el atributo más pequeño basado en una definición adecuada sobre la función que combine la clasificación exacta y el tamaño del conjunto de atributos [56]. El GWO provee buenos tiempos de respuesta, como también lo hace de la misma al entregar una cantidad de datos importantes en un tiempo realmente efectivo. Además, el algoritmo se aproxima a brindar muchos datos robustos, comparado con otro tipo de algoritmos fusionados con RST, otra mejora importante desarrollada en un artículo publicado en 2015 es implicar algunas características de selección y diferentes funciones que en el futuro se espera obtener mejores resultados [57].

3.6.2 TENDENCIA DE RST Y ARTIFICIAL FISH SWARM

El algoritmo de AFSA que proviene de sus siglas en inglés Artificial Fish Swarm Algorithm, está basado en la optimización que simula el comportamiento de un cardumen de peces, siguiendo con

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

búsquedas locales para obtener un óptimo global, el cual sea exitoso aplicado a solucionar diferentes problemas de combinación. Este es un estocástico y paralelamente es un algoritmo de búsqueda. Además, no necesita saber la información concreta de un problema, en vez de eso solo necesita comparar las ventajas y desventajas de la solución de dichos problemas [58]. El AFSA es un proceso que propone encontrar la reducción mínima de una manera eficiente para observar el cambio del significado del subconjunto de atributos y el número de experimentos de atributos han demostrado que proponer este algoritmo ha logrado resultados de manera más eficiente que otros métodos, generando menos reglas, esto significa que las reglas son más generales y suelen ser mejores para reconocer nuevos objetos, tales como reducciones que puede ser más aplicadas a prácticas en máquinas de aprendizaje [59].

3.6.3 TENDENCIA DE RST Y ALGORITMO DE LUCIERNAGA

El algoritmo de la luciérnaga (FA de sus siglas en inglés Firefly Algorithm) el cual simula el sistema de atracción real de estos insectos, los cuales producen unos flashes luminosos como una señal de un sistema de comunicación con otras luciérnagas, especialmente para llamar la atención. El FA expresa esta característica de este flash como un objetivo de un problema que debe ser optimizado. Además, el algoritmo no necesita ningún parámetro de suposición al azar y

produce el mismo resultado cada vez, este documento expresa una nueva característica de selección del algoritmo, que incorpora el comportamiento básico de las luciérnagas fusionándolo con RST para mejorar los tiempos de respuesta, estos tiempos del algoritmo propuesto fueron evaluados comparándolos con otros métodos usando conjuntos de datos médicos [60]. El atributo de reducción es usado para permitir la eliminación de atributos redundantes, sin dejar de ser toda la ambigüedad restante de los conjuntos de datos originales, este documento propone una estrategia de atributo de reducción óptimos, apoyándose con RST y FA discretos. Para demostrar la aplicabilidad y la superioridad del modelo propuesto, comparado entre modelos propuestos populares que también han sido investigados [61].

3.6.4 TENDENCIA DE RST Y POLINIZACION DE FLOR

El algoritmo de polinización de flor (FPA de sus siglas en inglés Flower Pollination Algorithm) es una nueva técnica inspirada en el proceso de polinización de flores, en este documento se propone una selección de características basadas en RST con FPA, el cual es típicamente asociado con la transferencia de polen que realizan los insectos, pájaros, murciélagos y otros animales, que puede realizarse de una flor de diferentes plantas, mientras que la polinización propia es una fertilización de una sola flor. El uso de este método provee una mejora en la búsqueda de

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

espacio en los atributos y la convergencia de soluciones óptimas. También provee mejor robustez contra diferentes inicializaciones al azar [62].

y así poderse aplicar e implementar en diferentes procesos de optimización como se observó en el presente documento.

CONCLUSIONES

Se expone en este documento por medio de una agrupación de los trabajos publicados por diferentes autores, que se han venido realizando en el campo de RST y sus diferentes aplicaciones y metodologías, tipos y fusiones, todo enfocado a la mejora del proceso en sí para la obtención de resultados, atributos o tiempos de respuesta. Se ha podido identificar que es una metodología usada en diferentes aplicaciones, implementada de diferentes maneras y lo suficientemente variable para aplicarse a cualquier proceso, no solo para la minería de datos sino también para asuntos específicos, siendo una herramienta útil para la toma de decisiones, convirtiéndose en un tópico que está atrayendo la atención de los investigadores. Se demostró que los algoritmos PSO siguen siendo los más utilizados, que las nuevas metodologías, donde llama la atención que son fusiones con algoritmos inspirados en el comportamiento de la naturaleza, especialmente de los animales, se está convirtiendo en tendencia, ampliando así el campo de acción e investigación del área que se mantendrá en los años venideros. A su vez se resalta la capacidad de la teoría y su versatilidad para combinarse con diferentes metodologías

REFERENCIAS

- [1]. Z. Pawlak. "Rough Sets". International journal of Computer and Information Sciences. Vol. 11. 1982. pp. 341-356.
- [2]. Luisa Fernanda Múnera-Monsalve, Maricela Marulanda-Usma, Edwin Alejandro Pérez-Ramírez, Delio Augusto Aristizábal-Martínez, Jorge Alberto Jaramillo-Garzón. (2013) Conjuntos Aproximados Y Algoritmos Genéticos Para El Análisis De La Deserción Universitaria Con El Toolbox TurboGAEn Matlab. Colección de Resúmenes I Encuentro de Tecnología e Ingeniería y IX Simposio Internacional en Energías, 37-39
- [3]. Filiberto, Y., Bello, R., Caballero, Y., & Larrua, R. (2011). A measure in the rough set theory to decision systems with continuo features. Revista de la Facultad de Ingeniería de la Universidad Antioquia, No. 60, pp. 141-152.
- [4]. Zhang, J., Li, T., Ruan, D., Gao, Z., & Zhao, C. (2012). A parallel method for computing rough set approximations. Information Sciences, 194, 209-223.
- [5]. Zhang, J., Li, T., Ruan, D., & Liu, D. (2012). Neighborhood rough sets for dynamic data mining. International

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Journal of Intelligent Systems, 27(4), 317-342.

[6]. Jia, X., Tang, Z., Liao, W., & Shang, L. (2014). On an optimization representation of decision-theoretic rough set model. *International Journal of Approximate Reasoning*, 55(1), 156-166.

[7]. Qian, Y., Liang, J., Pedrycz, W., & Dang, C. (2010). Positive approximation: an accelerator for attribute reduction in rough set theory. *Artificial Intelligence*, 174(9), 597-618.

[8]. Xie, F., Lin, Y., & Ren, W. (2011). Optimizing model for land use/land cover retrieval from remote sensing imagery based on variable precision rough sets. *Ecological Modelling*, 222(2), 232-240.

[9]. Jia, X., Liao, W., Tang, Z., & Shang, L. (2013). Minimum cost attribute reduction in decision-theoretic rough set models. *Information Sciences*, 219, 151-167.

[10]. D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. General Syst.*, vol. 17, pp. 191–209, 1990.

[11]. Molodtsov, D. (1999). Soft set theory—first results. *Computers & Mathematics with Applications*, 37(4), 19-31.

[12]. Suguna, N., & Thanushkodi, K. G. (2011). An independent rough set approach hybrid with artificial bee colony algorithm for dimensionality

reduction. *American Journal of Applied Sciences*, 8(3), 261.

[13]. Chen, J., & Zhang, C. (2011). Efficient clustering method based on rough set and genetic algorithm. *Procedia Engineering*, 15, 1498-1503.

[14]. Cheng, C. H., Chen, T. L., & Wei, L. Y. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Information Sciences*, 180(9), 1610-1629.

[15]. Delio Augusto Aristizabal Martinez, Luisa Fernanda Munera Monsalve, Jorge Alberto Jaramillo Garzon (2012) *Predicción de la Deserción Universitaria utilizando Rough Sets y algoritmos genéricos con Rosseta System*. Libro de resúmenes Tercer Congreso de Inteligencia Computacional, 11-14.

[16]. Lin, C. H. (2013). A rough penalty genetic algorithm for constrained optimization. *Information Sciences*, 241, 119-137.

[17]. Derrac, J., Cornelis, C., García, S., & Herrera, F. (2012). Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. *Information Sciences*, 186(1), 73-92.

[18]. Nair, B. B., Mohandas, V. P., & Sakthivel, N. R. (2010). A decision tree—rough set hybrid system for stock market trend prediction. *International Journal of Computer Applications*, 6(9), 1-6.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

- [19]. Zhou, L., & Jiang, F. (2011). A rough set based decision tree algorithm and its application in intrusion detection. In Pattern Recognition and Machine Intelligence (pp. 333-338). Springer Berlin Heidelberg.
- [20]. Lei, S. H. I., Weng, M., Xinming, M. A., & Lei, X. I. (2010). Rough set based decision tree ensemble algorithm for text classification. *Journal of Computational Information Systems*6, 1, 89-95.
- [21]. Sharma, R., Jain, P., & Shrivastava, S. K. (2012). An Optimize Decision Tree Algorithm Based on Variable Precision Rough Set Theory Using Degree of β -dependency and Significance of Attributes. *IJCSIT) International Journal of Computer Science and Information Technologies*, 3(3), 3942-3947.
- [22]. Chen, Y., Miao, D., Wang, R., & Wu, K. (2011). A rough set approach to feature selection based on power set tree. *Knowledge-Based Systems*, 24(2), 275-281.
- [23]. Yamany, W., El-Bendary, N., Zawbaa, H. M., Hassanien, A. E., & Snášel, V. (2014). Rough Power Set Tree for Feature Selection and Classification: Case Study on MRI Brain Tumor. In *Innovations in Bio-inspired Computing and Applications* (pp. 259-270). Springer International Publishing.
- [24]. Zhai, J. H. (2011). Fuzzy decision tree based on fuzzy-rough technique. *Soft Computing*, 15(6), 1087-1096.
- [25]. bin Basir, M. A., & binti Ahmad, F. Comparison on Swarm Algorithms for Feature Selections/Reductions.
- [26]. Rana, S., Jasola, S., & Kumar, R. (2011). A review on particle swarm optimization algorithms and their applications to data clustering. *Artificial Intelligence Review*, 35(3), 211-222.
- [27]. Bae, C., Yeh, W. C., Chung, Y. Y., & Liu, S. L. (2010). Feature selection with intelligent dynamic swarm and rough set. *Expert Systems with Applications*,37(10), 7026-7032.
- [28]. Zhao, Q., Meng, G., Zhao, J., & Liu, L. (2010, April). A new method of data mining based on rough sets and discrete particle swarm optimization. In *Computer Engineering and Technology (ICCET), 2010 2nd International Conference on* (Vol. 2, pp. V2-424). IEEE.
- [29]. Xue, B., Cervante, L., Shang, L., Browne, W. N., & Zhang, M. (2014). Binary PSO and rough set theory for feature selection: A multi-objective filter based approach. *International Journal of Computational Intelligence and Applications*,13(02), 1450009.
- [30]. Adamczyk, M. (2014, September). Parallel feature selection algorithm based on rough sets and particle swarm optimization. In *Computer Science and Information Systems (FedCSIS), 2014 Federated Conference on* (pp. 43-50). IEEE.
- [31]. Yue, B., Yao, W., Abraham, A., & Liu, H. (2007, June). A new rough set reduct algorithm based on particle swarm

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

optimization. In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 397-406). Springer Berlin Heidelberg.

[32]. Chebrolu, S., & Sanjeevi, S. G. (2015). Attribute Reduction in Decision-Theoretic Rough Set Model using Particle Swarm Optimization with the Threshold Parameters Determined using LMS Training Rule. *Procedia Computer Science*, 57, 527-536.

[33]. Bakar, A. A., Hamdan, A. R., & Nazri, M. Z. A. (2010, November). A rough set outlier detection based on Particle Swarm Optimization. In 2010 10th International Conference on Intelligent Systems Design and Applications (pp. 1021-1025). IEEE.

[34]. Miao, A., Shi, X., Zhang, J., Jiang, W., Zhang, J., & Gui, X. (2010, March). Dynamic particle swarm optimization based on neighborhood rough set model. In *Informatics in Control, Automation and Robotics (CAR), 2010 2nd International Asia Conference on* (Vol. 2, pp. 95-100). IEEE.

[35]. Deng, W., Yang, X., Zou, L., Zhao, H., & Li, W. (2012). An efficient fusion approach to rule extraction based on rough set theory and particle swarm optimization and its application. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 226(7), 904-913.

[36]. Moaref, A., & Naeini, V. S. (2013, August). A particle swarm optimization

based on a ring topology for fuzzy-rough feature selection. In *Fuzzy Systems (IFSC), 2013 13th Iranian Conference on* (pp. 1-6). IEEE.

[37]. Al Daoud, E. (2015). An Efficient Algorithm for Finding a Fuzzy Rough Set Reduct Using an Improved Harmony Search. *a A*, 5, 2.

[38]. Jensen, R., & Cornelis, C. (2011). Fuzzy-rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42), 5871-5884.

[39]. Chen, D., Zhang, L., Zhao, S., Hu, Q., & Zhu, P. (2012). A novel algorithm for finding reducts with fuzzy rough sets. *Fuzzy Systems, IEEE Transactions on*, 20(2), 385-389.

[40]. Çağman, N., & Enginoğlu, S. (2010). Soft set theory and uni-int decision making. *European Journal of Operational Research*, 207(2), 848-855.

[41]. Xiao, Z., Chen, W., & Li, L. (2013). A method based on interval-valued fuzzy soft set for multi-attribute group decision-making problems under uncertain environment. *Knowledge and information systems*, 34(3), 653-669.

[42]. Xin, G., Qiang, G., Jing, Z., & Zhengchao, Z. (2010, October). An attribute reduction algorithm based on rough set, information entropy and ant colony optimization. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on* (pp. 1313-1317). IEEE.

[43]. Wu, J., Qiu, T., Wang, L., & Huang, H. (2011). An Approach to Feature

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

Selection Based on Ant Colony Optimization and Rough Set. In Intelligent Computing and Information Science (pp. 466-471). Springer Berlin Heidelberg.

[44]. Chebrolu, S., & Sanjeevi, S. G. (2015, August). Attribute Reduction on Continuous Data in Rough Set Theory using Ant Colony Optimization Metaheuristic. In Proceedings of the Third International Symposium on Women in Computing and Informatics (pp. 17-24). ACM.

[45]. Chen, Y., Miao, D., & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. Pattern Recognition Letters, 31(3), 226-233.

[46]. Mohan, B. C., & Baskaran, R. (2012). A survey: Ant Colony Optimization based recent research and implementation on several engineering domain. Expert Systems with Applications, 39(4), 4618-4627.

[47]. Changdar, C., Pal, R. K., & Mahapatra, G. S. (2016). A genetic ant colony optimization based algorithm for solid multiple travelling salesmen problem in fuzzy rough environment. Soft Computing, 1-15.

[48]. Vieira, S. M., Sousa, J. M., & Kaymak, U. (2012). Fuzzy criteria for feature selection. Fuzzy [48]. Sets and Systems, 189(1), 1-18.

[49]. Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial

bee colony (ABC) algorithm. Journal of global optimization, 39(3), 459-471.

[50]. Ye, D., & Chen, Z. (2015). A new approach to minimum attribute reduction based on discrete artificial bee colony. Soft Computing, 19(7), 1893-1903.

[51]. Suguna, N., & Thanushkodi, K. (2010). A novel rough set reduct algorithm for medical domain based on bee colony optimization. arXiv preprint arXiv:1006.4540.

[52]. Schiezero, M., & Pedrini, H. (2013). Data feature selection based on Artificial Bee Colony algorithm. EURASIP Journal on Image and Video Processing, 2013(1), 1.

[53]. Othman, Z., Bakar, A. A., Abdullah, S., Zakree, M., Nazri, M. Z. A., & Sengalang, N. A. (2012). Dynamic tabu search for dimensionality reduction in rough set. WSEAS Transactions on Computers, 11(4), 89-100.

[54]. Wang, J., Guo, K., & Wang, S. (2010). Rough set and Tabu search based feature selection for credit scoring. Procedia Computer Science, 1(1), 2425-2432.

[55]. Banerjee, S., & Al-Qaheri, H. (2011). An intelligent hybrid scheme for optimizing parking space: A Tabu metaphor and rough set based approach. Egyptian Informatics Journal, 12(1), 9-17.

	INFORME FINAL DE TRABAJO DE GRADO	Código	FDE 089
		Versión	03
		Fecha	2015-01-22

[56]. Yamany, W., Emary, E., & Hassanien, A. E. (2016). New Rough Set Attribute Reduction Algorithm Based on Grey Wolf Optimization. In The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015), November 28-30, 2015, Beni Suef, Egypt (pp. 241-251). Springer International Publishing.

[57]. Emary, E., Zawbaa, H. M., Grosan, C., & Hassenian, A. E. (2015). Feature subset selection approach by gray-wolf optimization. In Afro-European Conference for Industrial Advancement (pp. 1-13). Springer International Publishing.

[58]. Wang, F., Xu, J., & Li, L. (2014, October). A novel rough set reduct algorithm to feature selection based on artificial fish swarm algorithm. In International Conference in Swarm Intelligence (pp. 24-33). Springer International Publishing.

[59]. Chen, Y., Zhu, Q., & Xu, H. (2015). Finding rough set reducts with fish swarm algorithm. Knowledge-Based Systems, 81, 22-29.

[60]. Banati, H., & Bajaj, M. (2011). Fire fly based feature selection approach. IJCSI International Journal of Computer Science Issues, 8(4).

[61]. Long, N. C., Meesad, P., & Unger, H. (2014). Attribute reduction based on rough sets and the discrete firefly algorithm. In Recent Advances in Information and Communication Technology (pp. 13-22). Springer International Publishing.

[62]. Zawbaa, H. M., Hassanien, A. E., Emary, E., Yamany, W., & Parv, B. (2015, December). Hybrid flower pollination algorithm with rough sets for feature selection. In 2015 11th International Computer Engineering Conference (ICENCO) (pp. 278-283). IEEE.



INFORME FINAL DE
TRABAJO DE GRADO

Código	FDE 089
Versión	03
Fecha	2015-01-27

FIRMA ESTUDIANTES

Esteban Salazar

2222222

FIRMA ASESOR

[Signature]

FECHA ENTREGA:

ENERO 27/17

*Revisión sobre
diferentes métodos
de optimización
basados en
rough set theory*

FIRMA COMITÉ TRABAJO DE GRADO DE LA FACULTAD _____

RECHAZADO _____

ACEPTADO _____

ACEPTADO CON MODIFICACIONES _____

ACTA NO. _____

FECHA ENTREGA: _____

FIRMA CONSEJO DE FACULTAD _____

ACTA NO. _____

FECHA ENTREGA: _____